Fayoum University Faculty of Engineering, 2024, Vol: 7(1), 130-141
https://fuje.journals.ekb.eg/
ISSN Online: 2537-0634
ISSN Print: 2537-0626

Instance Segmentation and Classification of Coffee Leaf Plant using Mask RCNN and Transfer Learning

Fatma M. A. Mazen [1], Ahmed A. Nashat [1,*]

[1]Electrical Engineering Department, Faculty of Engineering, Fayoum University, Fayoum 63514, Egypt

*Corresponding author: Ahmed A. Nashat (aan01@ fayoum.edu.eg).

How to cite this paper: Mazen, F.M.A., and Nashat, A.A. (2024). Instance Segmentation and Classification of Coffee Leaf Plant using Mask RCNN and Transfer Learning, *Fayoum University Journal of Engineering, Vol: 7(1), 130-141*
https://dx.doi.org/10.21608/FUJE.2023.226247.1057

Abstract

Coffee is one of the most consumed beverages in the world. It is crucial in the economy of many industrial companies in developing countries. This study proposes a deep learning algorithm called Mask RCNN to segment coffee leaves from complex real-world backgrounds and classify them as healthy and unhealthy. The RoCole dataset was manually labeled using the VGG Image annotator. The algorithm uses Resnet101 and the FPN architecture for feature extraction. The RPN creates region proposals for each feature map to separate the input image from the background. The system has a high-test accuracy of 97.76% for the binary classifier. If the image is classified as unhealthy, it goes through another segmentation stage based on the HSV color model to highlight the defective areas of the coffee leaf. The instance segmentation results showed that the mAP@50:95 was 100%, the recall@50:95 was 84.5%, and the F1-score was 91.6%.

Keywords

Convolutional Neural Networks; Instance Segmentation and classification; Coffee Leaf; RoCole; Plant Diseases; VIA

1. Introduction

Coffee is the most important agricultural product in international trade. It is a major commodity on the global market and provides a source of revenue for millions of people concerned with the crop's cultivation, marketing, and processing. More than 50 developing countries in Africa, Asia, and South America are involved in coffee production and exportation to consumers in the industrialized countries. Coffee leaf Plant diseases are one of the main factors to decrease the yield and the quality of global agriculture production. Some of the main problems in coffee crops that affect the coffee tree, causing losses and reducing productivity, are pests, bacteria, fungi, nematodes, and viruses. Other causes are lack of essential nutrients for the growth of plants, water, and light. Early identification, classification, and localization of coffee leaf Plant diseases reduce the chance of economic losses with the production. To ensure coffee products' competitivity, reliability, quality standards, and productivity, automatic image processing tools based upon intelligent techniques are paramount over visual features methods. Such a machine vision system can make decisions at a high-speed rate. This paper proposes a method for detecting and classifying the affected parts of the coffee leaf using image segmentation and classification techniques. First, mask RCNN is used during the instance segmentation phase. Second, a transfer learning VGG16 pre-trained model is used for the classification stage to classify the segmented image into healthy or unhealthy leaves. Finally, an HSV-based segmentation is applied to the unhealthy image to detect the diseased parts.

2. Related Research

Many methods, which depend on color, texture features, classical machine learning, and deep learning, have been developed to classify coffee leaf diseases. Yebasse et al. (2021) used Visualization techniques like Grad-CAM, Grad-CAM++, and Score-CAM to highlight significant areas responsible for the model's classification. In their study, they demonstrate the relevance of visualization in coffee disease classification. In support of their argument, they present the naïve approach and the guided approach, using the Residual U-block (RSU) and the nested U-structure for object detection (2020) and compare their accuracy and visualization results. The guided approach achieved a classification accuracy of 98%, while the naïve approach achieved an accuracy of 77% on the Robusta coffee leaf image dataset (RoCole). They developed a practical method to localize coffee diseases without having to use annotating images with object bounding boxes during training.

Manso et al. (2019) developed a mobile application to detect and classify coffee leaf rust and miner diseases. They tested the k-means algorithm, the Otsu method, and the iterative threshold method after transforming the input image to YCbCr color space for the segmentation stage. They calculated attributes of color and texture for feature extraction. Finally, they used Artificial Neural Network trained with Backpropagation and Extreme Learning Machine for the classification stage. They achieved a classification accuracy of 96.013%.

Valliammal, N., Geethalakshmi, S. N (2012), a 3-stage segmentation approach for plant leaf images using fuzzy clustering is proposed. In the first stage, the input image was transformed to grayscale and then passed through a median filter to reduce noise. Secondly, some morphological operations like opening and closing was applied to smoothen the image. Then, Wavelet transform in the third stage was utilized to get the leaf boundaries. Finally, a Fuzzy threshold was employed to get the exact shape and edges of the leaf.

Gao and Lin (2019) presented a fully automatic and accurate segmentation approach for medicinal plant leaf images in complex background. First, they tried OTSU segmentation on different component images. Next, the image gradient was utilized to get an accurate foreground by enhancing and extracting the veins. Finally, they used the marker-controlled watershed approach to realize image segmentation. They conducted several experiments on a widely used database and another self-built database.

Convolutional neural networks (CNN) and transfer learning are a class of deep learning neural networks and have acquired a broad application in image classification. They are powerful visual models that yield hierarchies of features. In research by Kumar et al. (2020), a transfer learning platform using the Inception v3 pre-trained model for detecting and classifying healthy and defected coffee plants is presented. They used data augmentation to increase the number of images in the training set. As a result, they achieved a classification accuracy of 97.61%.

In another research, Esgario et al. (2021) developed an Android platform application to help farmers identify coffee leaf diseases and pests. They utilized two of the state-of-the-art semantic segmentation architectures for the segmentation stage, UNet, Ronneberger et al (2015), and

PSPNet, Zhao et al (2017). They developed a framework to integrate the segmentation and classification tasks. In the segmentation stage, the model can segment each disease from the leaf and feed it to the classification network to classify each one. They achieved a mean Intersection over Union of 94.85% for semantic segmentation using UNet. As a result, they reached a classification accuracy of 97%.

Hsiao et al. (2014) conducted a comparative study on different techniques used for leaf image recognition. They proposed a novel learning-based leaf image recognition framework for automatic plant identification, where they formulated a leaf identification problem as a sparse representation problem. It encompasses two stages, the learning stage and the recognition stage. Sparse representation techniques efficiently solved various computer vision problems like action recognition and face recognition.

In a recent study, Esgario et al. (2019) used deep learning approach to classify and estimate the severity of coffee leaf biotic stress. They developed a new dataset of coffee leaf images to accomplish their experiments. In addition, they used data augmentation techniques to enlarge their dataset and enhance system robustness. Using ResNet50 architecture, they reached an accuracy of 86.51% for severity estimation and 95.24% for the biotic stress classification.

Liang et al. (2019) designed an efficient multitasking system called PD2 SE-Net to diagnose diseases, recognize the plant species and estimate the severity of diseases. They used the Plant Village dataset to conduct their experiments. They reached an overall accuracy of 98%, 99%, and 91% for plant disease classification, plant species recognition, and disease severity estimation, respectively.

Tassis et al. (2021) proposed an integrated framework by using different convolutional neural networks (CNN) to automate detection/recognition of lesions from in-field images collected via smartphone containing part of the coffee tree. In the first stage, they used a Mask R-CNN network for instance segmentation; in the second stage the UNet and PSPNet networks for semantic segmentation and finally, in the third stage, a ResNet for classification. For the Mask R-CNN network, they obtained a precision of 73.90% and a recall of 71.90% in the instance segmentation task. For the UNet and PSPNet networks, they obtained a mean intersection over union of 94.25% and 93.54%, respectively.

In their study, Jadhav S. B. et al. (2021) proposed an efficient soybean diseases identification method based on a transfer learning approach by using pretrained AlexNet and GoogleNet convolutional neural networks (CNNs). Their proposed AlexNet and GoogleNet CNNs were trained using over 1000 image samples of diseased and healthy soybean leaves to identify three soybean diseases. They used the five-fold cross-validation strategy. The proposed AlexNet and GoogleNet CNN-based models achieved an accuracy of 98.75% and 96.25%, respectively.

To classify diseases in bell pepper leaves, Bhagat, M. et al. (2023) employed three features such as local binary pattern (LBP) features, visual geometry group network (VGG-16) features and fused LBP & VGG-16 features. The accuracy obtained with random forest classifier for pepper bell dataset, where images captured in the farm, with LBP feature, VGG-16 feature and LBP + VGG-16 fused feature are 78.11%, 92.28% and 99.75% respectively.

Upadhyay, S. K. and Kumar, A. (2022), proposed an effective approach for recognition and identification of rice plant disease based on size, shape, and color of lesions in the leaf image. Their proposed model applies Otsu's global thresholding technique to perform image binarization to remove background noise of the image. The offered method based on fully connected CNN was trained using 4000 image samples of each diseased leaves and 4000 image samples of healthy rice leaves, to detect the three rice diseases. Results show that the nominated fully connected CNN is fast and effective approach, which provides an accuracy of 99.7% on the dataset.

Sachar, S., and Kumar, A. (2022) developed an Ensemble Deep Learning- Automatic Medicinal Leaf Identification (EDL-AMLI) classifier based on the weighted average of the pre-train neural networks MobileNetV2, InceptionV3, and ResNet50 outputs. They showed that the EDL-AMLI classifier outperformed the Dense Layer classifier of the state-of-the-art pre-trained MobileNetV2, InceptionV3, and ResNet50 models. Their model achieved 99.66% accuracy on the test set and average accuracy of 99.9% using threefold and fivefold cross validation.

In research by Fenu, G., Malloci, F. M. (2022), the issue of plant diseases classification was addressed. Two plant species and five biotic stresses are analyzed using different architectures, such as EfficientB0, MobileNetV2, InceptionV2, ResNet50 and VGG16. Experiments show that model performance drops drastically when using representative datasets. In detail, MobileNetV2 and EfficientNetB0 achieved very close performance for laboratory datasets and outperformed the other architectures. MobileNetV2 scored 87.02% (RoCoLe), 87.31% (Plant Pathology), while EfficientNetB0 scored 82.90% (RoCoLE) and 82.57% (Plant Pathology). VGG16 achieved high results for the BRACOL (96.54%) and Plant Village (97.84%) datasets but had a marked decline in accuracy for the field datasets, 79.83% for RoCoLe, and 70.11% for Plant Pathology. Lower results across all datasets were recorded by ResNet50.

The rest of the paper is organized as follows: Sect. 3 illustrates the dataset. Sect. 4 describes the proposed segmentation model. Sect. 5 demonstrates the classification system for distinguishing between the healthy and the unhealthy coffee leaf classes. In Sect. 6, the proposed classification system is tested, and the results are discussed. Finally, Sect. 7 offers the conclusion.

3. The Dataset

The Robusta coffee leaf images dataset (RoCoLe), Parraga-Alava, J. et al (2019), will be used in this paper. It has 1560 coffee leaf images, 791 for healthy leaves and 769 for unhealthy leaves. A smartphone camera took images in the same coffee plant field in natural conditions. In addition, images of healthy and infected leaves' upper and back sides were taken. **Figure 1** shows examples of dataset healthy and unhealthy leaves, respectively.

(a) (b)

Figure 1. samples from RoCole leaves (a) Healthy leaf and (b) Unhealthy leaf.

4. Segmentation Phase

For instance segmentation, the Mask RCNN model is proposed, considered one of state-of-the-art convolutional neural networks. Instance segmentation means it can figure out the location and the pixels belonging to the object (i.e., coffee leaf). **Figure 2** shows the general framework of the Mask RCNN architecture. It has three branches; the first one is used to output the class label, the second produces a bounding box for the candidate object, and the third branch produces the object mask (Region of Interest).

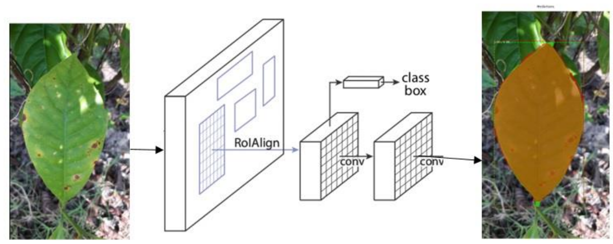


Figure 2. Framework of the proposed Mask RCNN, Johnson, J. W. (2018).

In order to perform instance segmentation, VGG Image Annotator (VIA), Dutta, A. et al (2019), has been used to separate coffee leaves from the complex background with polygon masks. VIA is a simple, standalone manual annotation software for images, audio, and video. **Figure 3** shows samples of the manually annotated images of the RoCole dataset using the VIA tool. The annotation file is in "JSON" format, which contains the coordinates of all the polygons manually drawn on the images. The training set (1412 images) and test set (148 images) have been labeled separately. The annotation file of the test set carries the exact name of the training set's annotation file.

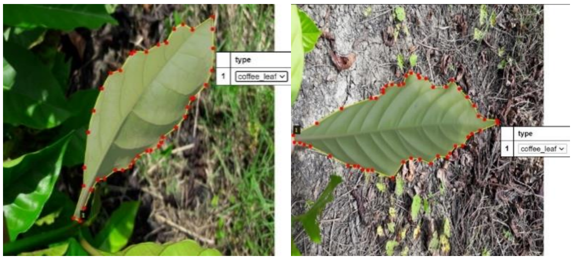


Figure 3. samples of manually annotated images.

Error! Reference source not found. **1** summarizes the parameter settings of Mask RCNN. In the original implementation of Mask RCNN, the mask shape is 28 x 28. When setting mask shape to this value, the masks have straight-edged cut-offs or stair-like edges, as shown in **Figure 4**.

Table 1. Parameter settings of Mask RCNN

Parameter	Value
STEPS_PER_EPOCH	100
VALIDATION_STEPS	50
BACKBONE	resnet101
RPN_ANCHOR_RATIOS	[0.5,1,2]
RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	128
POST_NMS_ROIS_TRAINING	2000
POST_NMS_ROIS_INFERENCE	1000
IMAGE_RESIZE_MODE	square
IMAGE_MIN_DIM	1024
IMAGE_MAX_DIM	1024
USE_MINI_MASK	True
MINI_MASK_SHAPE	(56,56)
MASK_SHAPE	[56, 56]
DETECTION_NMS_THRESHOLD	0.3
LEARNING_RATE	0.001
LEARNING_MOMENTUM	0.9
WEIGHT_DECAY	0.0001
DETECTION_MIN_CONFIDENCE	0.9
NUM_CLASSES	2
IMAGES_PER_GPU	1

Figure 4. 28 x 28 mask with straight edged cut offs and stair like edges.

For higher quality and smooth masks, we modified the design of the Feature Pyramid Network (FPN) by adding another Conv2DTranspose layer which acts as upsampling layer resulting in a mask shape of 56 x 56. RPN_TRAIN_ANCHORS_PER_IMAGE parameter has been decreased from 256 to 128 to avoid memory crashes. NUM_CLASSES has been set to 2 since we have two classes, the background class and the coffee_leaf class. USE_MINI_MASK parameter has been set to True due to high-resolution images of the Rocole dataset. The highest resolution is 2322 x 4128, and the lowest resolution is 720 x 1280. **Figure 5** illustrates the idea of using the mini mask. IMAGE_RESIZE_MODE has been set to square, which means that all images will be resized and padded with zeros to get a square image of size [IMAGE_MAX_DIM, IMAGE_MAX_DIM]. IMAGE_MAX_DIM has been set to 1024, which means that all images will be resized to 1024 x 1024.

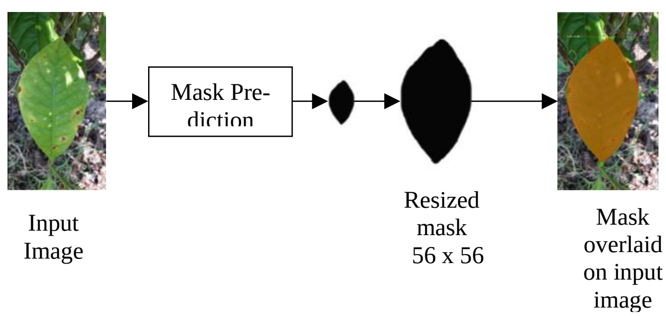


Figure 5. Idea of using mini mask in Mask RCNN.

5. Classification Phase

Traditional machine learning and deep learning algorithms are designed to work separately. Those models are trained for a specific task. Once the feature-space distribution changes, the model must be rebuilt from scratch. Transfer learning has conquered the isolated learning approach, and the knowledge acquired for one task is utilized to solve related ones, Zhuang, F. et al (2021). Moreover, using a pre-trained model, the same result can be accomplished using far fewer data. Training models using transfer learning improve their generalization ability and make them more accessible to non-experts in the deep learning field. We used the VGG16 model; a pre-trained model trained using the ImageNet dataset with 1000 classes.

The original VGG16 has two main parts, the convolutional base, and the classifier. The convolutional base has five convolutional blocks. The first and second blocks have two convolutional layers and one Maxpooling layer. On the other hand, the third, the fourth, and the fifth convolutional blocks have three convolutional layers and one Maxpooling layer. The convolutional base part is kept the same as the original VGG16 architecture.

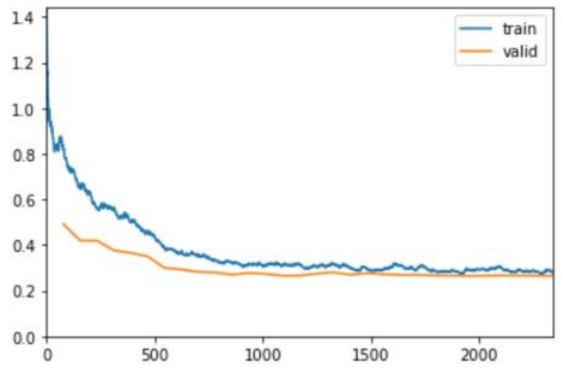
Furthermore, we added a flatten layer to change the dimensions of the tensor input of the previous layer and ensure that the output size is a 1x1 tensor with a length corresponding to the input tensor volume. The original classifier part is discarded. A new Fully connected layer of 512 neurons and relu activation function is added. Dropout, Nandini, G. S. et al (2021), is an effective and powerful regularization technique that aims to randomly drop out or ignore several neurons during the training phase. It is an excellent means to reduce overfitting, so a dropout layer with a dropout rate of 0.5 is added after the Fully connected layer. Finally, a fully connected layer of one neuron and sigmoid activation function is added since our problem is a binary classification problem.

6. Experimental Results and Discussion

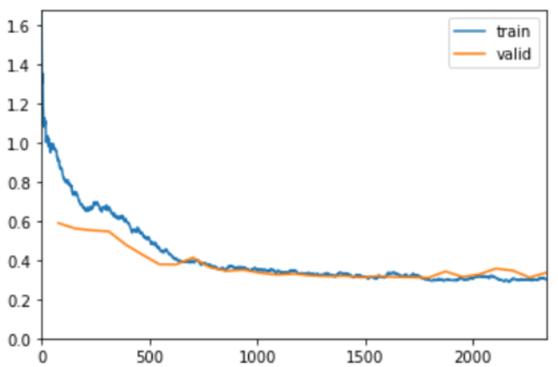
Figure 6 shows the output of different stages of the desired system. **Figure 6 (a)** shows samples of the input images to the mask RCNN model. **Figure 6 (b)** shows the output mask overlaid on the input image. Next, the output mask is bitwise ANDed with the input image producing the segmented image shown in **Figure 6 (c)**. The output of the segmentation stage is then passed to a binary classifier to determine whether the segmented image is healthy or unhealthy. If the output class of the binary classifier is unhealthy, the segmented image will later be passed through another segmentation stage based on the HSV color model to highlight the diseased parts of the coffee leaf. The result is shown in **Figure 6 (d)**.

Figure 6. Samples of the mask RCNN segmentation model results. (a) Input image, (b) mask overlaid on input image, (c) the segmented image, and (d) the diseased part.

For classification, all images have been resized to 256 x 256. The VGG16 model was then trained using center resized crops of 224 x 224 with a learning rate of 0.002290867641568184 for 30 epochs and a batch size of 16. The dataset is split the following way: 80% for training and validation, corresponding to 1248 images, 20% for the test, 151 for healthy leaves, and 161 for unhealthy leaves. To increase the number of images in the training and validation dataset, we have employed data augmentation techniques, Buslaev, A. et al (2020). Data augmentation incorporates modified copies of existing data to enhance the model's performance and generalizability. Examples are horizontal and vertical flipping, shift scale rotation, and random brightness contrast. To demonstrate the effect of the segmentation phase, we have trained two models, one using the controlled background Rocole dataset and another using the complex background version. **Figure 7** shows train and validation loss curves for such VGG16 two models.



(a)



(b)

Figure 7. VGG16 Model loss curves (a) controlled background and (b) complex background

The training loss and the validation loss curves decrease and stabilize at about the same error value and the same epoch. Therefore, the model accurately fits the training data. The training dataset is learned perfectly. Results indicate that no further training is needed. However, the learning curve for the complex background shows noisy movement of the validation loss around the training loss after 1750 iterations which correspond to 22 epochs. This noisy movement suggests that the validation dataset, for the complex background, has too few examples compared to the training dataset and does not provide sufficient information to evaluate the ability of the model to generalize after 22 epochs. Therefore, end training at 22 epochs for the complex background Rocole dataset is reasonably fair.

Sensitivity and precision are two statistical performance measures for classification tests. Sensitivity, also known as recall, is the prediction model's ability to select the instance of a particular class from the dataset. It is the proportion of the actual positive classes which are correctly identified. On the other hand, precision is defined as the proportion of the predicted positive classes which are correctly identified. Overall prediction accuracy is another useful statistical measure that describes the performance of a classifier. Overall prediction accuracy is the proportion of the total number of correct predictions. F1-score, another essential evaluation metric, measures a model's accuracy on a dataset. It is used to evaluate binary classification systems. It is the harmonic mean of the model's precision and recall.

Tables 2 and 3 show the confusion matrix of the classification model trained using the controlled background and the complex background Rocole dataset. The performance measures of the classifier for both two datasets in terms of the F1-score, precision, and recall metrics are shown in **Figure 8** Simulation results show that the class F1-score and the overall prediction accuracy are very close in value, about 97.76% for the controlled dataset and about 95.51% for the complex dataset. This result is expected since both datasets are balanced. In addition, results reveal that the segmentation phase has improved the overall model accuracy by 2.25%, which is an outstanding achievement.

The binary classifier using the two datasets shows high scores for both precision and recall. Thus, the model is perfect, and it returns accurate results. However, classification using the controlled background Rocole dataset is more precise and sensitive. It ranges from 96.15% to 99.34%. Whereas classification using the complex background Rocole dataset ranges from 93.63% to 97.42%.

Table 2. Confusion matrix of the classification model using the controlled background Rocole dataset

	Predicted Class		
Actual Class	Healthy	Unhealthy	Class Sensitivity (Recall) %
Healthy	150	1	99.34
Unhealthy	6	155	96.27
Class Precision %	96.15	99.36	Overall Prediction Accuracy =
Class F1-score %	97.72	97.79	97.76%

Table 3. Confusion matrix of the classification model using the complex background Rocole dataset

	Predicted Class		
Actual Class	Healthy	Unhealthy	Class Sensitivity (Recall) %
Healthy	147	4	97.35
Unhealthy	10	151	93.79
Class Precision %	93.63	97.42	Overall Prediction Accuracy =
Class F1-score %	95.45	95.57	95.51%

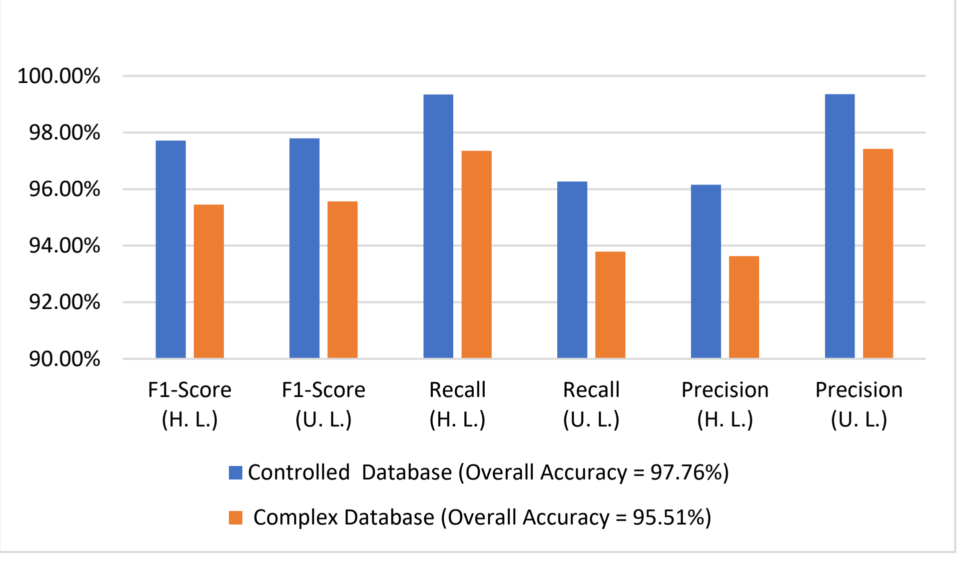


Figure 8. Performance measure of the classifier for the controlled and the complex dataset
Notes: H. L. stands for healthy leaf, and U. L. stands for unhealthy leaf

Some correct and misclassified images of the controlled and the complex background Rocole dataset are shown in **tables 4 and 5**, respectively. The sun's rays sneaking through the trees and falling over parts of the leaves yielding bright spots are a typical pattern for misclassification.

Table 4. Samples of the correct classification model results. (a) The controlled background Rocole dataset. (b) The complex background Rocole dataset

				
Actual Image	Healthy	Healthy	Unhealthy	Unhealthy
Predicted Image	Healthy	Healthy	Unhealthy	Unhealthy

(a)

				
Actual Image	Healthy	Healthy	Unhealthy	Unhealthy
Predicted Image	Healthy	Healthy	Unhealthy	Unhealthy

(b)

Table 5. Samples of the misclassified classification model results. (a) The controlled background Rocole dataset. (b) The complex background Rocole dataset

				
Actual Image	Unhealthy	Unhealthy	Unhealthy	Healthy
Predicted Image	Healthy	Healthy	Healthy	Unhealthy

(a)

				
Actual Image	Healthy	Healthy	Healthy	Unhealthy
Predicted Image	Unhealthy	Unhealthy	Unhealthy	Healthy

(b)

Results of the proposed modified Mask RCNN model are compared to previous studies conducted using the "RoCole" dataset, Fenu, G., Malloci, F. M. (2022). **Table 6** shows the accuracy, precision, and F1-score of our proposed modified Mask RCNN model, EfficientB0, MobileNetV2, InceptionV2, ResNet50 and VGG16. It is seen that our proposed model has outperformed the other five models. It has achieved at least 10.74% increase in accuracy over the other five models. Evaluating the models using the precision, and F1-score metrics confirms the rankings identified with the two metrics, in which our proposed modified Mask RCNN performed best, followed by MobileNetV2.

Table 6. Comparison of various studies conducted using the "RoCole" dataset, Fenu, G., Malloci, F. M. (2022)

Model	Accuracy	Precision	F1-Score
EfficientNetB0	82.90%	85.91%	82.18%
InceptionV2	85.36%	87.27%	85.16%
MobileNetV2	87.02%	89.23%	86.83%
ResNet50	67.40%	72.47%	85.45%
VGG16	79.83%	83.38%	79.28%
Our proposed modified Mask RCNN model	**97.76%**	**100%**	**91.60%**

7. Conclusion and Future Work

This paper presents a three-stage system: segmentation, classification, and disease localization of the coffee leaf. First, we developed a controlled background version of the RoCole dataset. Then, we created ground truth masks for the RoCole dataset using the VIA tool to further use in coffee leaf segmentation tasks. Mask RCNN is used during the instance segmentation phase. In contrast, transfer learning with the VGG16 pre-trained model is used for the classification stage to classify the segmented image into healthy or unhealthy leaves. The binary classifier achieved a test accuracy of 97.76%.
Further, an HSV-based segmentation is applied to the unhealthy image to detect the diseased parts. We utilize the mini mask idea to adopt high-resolution images during the instance segmentation phase. This paper also investigates the effect of the mini mask size on producing high-quality masks. We have modified the design of Mask RCNN to produce a mini mask of 56 x 56 instead of 28 x 28 to get rid of the straight-edged cut-offs and stair-like edges. The proposed models show superior performance in all stages. The instance segmentation model achieved a mean average precision of 100% at Intersection over Union of [0.5:0.95] and an F1-score of 91.6%. In the future, we aim to develop a mobile application to help farmers classify coffee leaf plants into healthy and unhealthy and localize the diseased part.
Future research lines will focus on investigation of other image segmentation and classification algorithms and on deploying the best model as a mobile application to assist untrained workers in recognizing healthy coffee leaf plants. Moreover, further study will focus on using other benchmarked datasets to analyze the behavior of the proposed model.

Statements and Declarations

Funding: No funding for this study.
Conflict of Interest: Authors have not received research grants from any company. Authors have not received a speaker honorarium from any company. Authors do not own stock in any Company. Authors are not a member of any committee. Authors declare that they

have no conflict of interest. Authors have no competing interests to declare that are relevant to the content of this article.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon request.

References

Abadi, M. & Agarwal et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. *arXiv*:1603.04467. https://doi.org/10.48550/arXiv.1603.04467

Bhagat, M., Kumar, D. & Kumar, S. (2023). Bell pepper leaf disease classification with LBP and VGG-16 based fused features and RF classifier. *International Journal of Information Technology*. **15**, 465–475. https://doi.org/10.1007/s41870-022-01136-z

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M. & Kalinin, A. A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*. **11**(2), 125. doi:10.3390/info11020125

Dutta, A., Guptam A. & Zissermann, A. (2019). VGG image annotator (VIA). *In Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, October 21–25, 2019, Nice, France. ACM, New York, NY, USA. https://doi.org/10.1145/3343031.3350535

Esgario, J. G. M, Castro, P. B. C., Tassis, L. M. & Krohling, R. A. (2021). An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. *Information Processing in Agriculture*. **9**(1), 38-47. doi: 10.1016/J.INPA.2021.01.004

Esgario, J. G. M., Krohling, R. A. & Ventura, J. A. (2019). Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*. 169. arXiv:1907.11561v1. doi.org/10.1016/j.compag.2019.105162

Fenu, G. & Malloci, F. M. (2022). Evaluating Impacts between Laboratory and Field-Collected Datasets for Plant Disease Classification. *Agronomy*. 12(10), 2359.. doi.org/10.3390/agronomy12102359

Gao, L. & Lin, X. (2019). Fully automatic segmentation method for medicinal plant leaf images in complex background. *Computers and electronics in agriculture*. **164**(5). doi: 10.1016/j.compag.2019.104924

Hsiao, J. K., Kang L. W., Chang, C. L. & Lin, C. Y. (2014). Comparative study of leaf image recognition with a novel learning-based approach. *In 2014 Science and Information Conference. IEEE*. 389-393. doi: 10.1109/SAI.2014.6918216

Jadhav, S. B., Udupi, V. R. & Patil, S. B. (2021). Identification of plant diseases using convolutional neural networks. *International Journal of Information Technology*. **13**, 2461–2470. https://doi.org/10.1007/s41870-020-00437-5.

Johnson, J. W. (2018). Adapting Mask-RCNN for Automatic Nucleus Segmentation. Machine Learning. *arXiv*:1805.00500. doi: 10.1007/978-3-030-17798-0

Kumar, M., Gupta, P. & Madhav, P (2020). Disease detection in coffee plants using convolutional neural network. *In 2020 5th International Conference on Communication and Electronics Systems (ICCES) IEEE*. 755-760. doi:10.1109/ICCES48766.2020.9138000

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D. & Sun, W. (2019). PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Computers and electronics in agriculture* 157. 518-529. doi: 10.1016/j.compag.2019.01.034

Manso, G. L., Knidel, H., Krohling, R. A. & Ventura. J. A. (2019). A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. *arXiv preprint* arXiv:1904.00742. https://doi.org/10.48550/arXiv.1904.00742

Nandini, G. S., Kumar, A. P. S. & Chidananda, K. (2021). Dropout technique for image classification based on extreme learning machine. *Global Transitions Proceedings*. **2**(1). 111-116. https://doi.org/10.1016/j.gltp.2021.01.015

Parraga-Alava, J., Cusme, K., Loor, A. & Santander, E. (2019). RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition. *Data in brief*. **25**. doi: 10.1016/j.dib.2019.104414

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R. & Jagersand, M. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition*. 106. doi: 10.1016/j.patcog.2020.107404

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Munich, Germany. 234-241. doi.org/10.1007/978-3-319-24574-4_28

Sachar, S. & Kumar, A. (2022). Deep ensemble learning for automatic medicinal leaf identification. *International Journal of Information Technology*. **14**, 3089–3097. https://doi.org/10.1007/s41870-022-01055-z

Tassis, L., Tozzi de Souza, J. & Krohling, R. (2021). A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Computers and Electronics in Agriculture*. Volume 186, 106191.

Upadhyay, S.K. & Kumar, A. (2022). A novel approach for rice plant diseases classification with deep convolutional neural network. *International Journal of Information Technology*. **14**, 185–199. https://doi.org/10.1007/s41870-021-00817-5

Valliammal, N. & Geethalakshmi, S. N. (2012). A novel approach for plant leaf image segmentation using fuzzy clustering. *International Journal of Computer Applications*. **44**(3), 10-20. doi:10.5120/6322-8669

Yebasse, M., Shimelis, B., Warku, H., Ko, J. & Cheoi, J. K. (2021).

Coffee Disease Visualization and Classification. *Plants*. **10**(6). doi: 10.3390/plants10061257

Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017). Pyramid scene parsing network. *Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*. Hawaii, USA. 2881-2890. doi: 10.1109/CVPR.2017.660

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y. & Zhu, H. (2021). A Comprehensive Survey on Transfer Learning. in *Proceedings of the IEEE*. **109**(1). 43-76. doi: 10.1109/JPROC.2020.3004555.